Diabetes Mellitus Abnormality Detection using AHMM (Auto Regressive Hidden Markov Model)

Dhivya. S

Assistant Professor, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam-636401.

Received 7 June 2016; Accepted 12 October 2016; Available 20 October 2016

ABSTRACT

Diabetes is a life-threatening issue in modern health care domain. With the use of data mining techniques, diabetes factors and co morbid conditions associated with diabetes has been found. In order to stifle the evolution of diabetes mellitus, applies distributed association rule mining and summarization techniques to electronic medical records. This helps to discover set of risk factors and co morbid conditions in distributed medical dataset using frequent itemset mining. In general, association rule mining (ARM) generates bulky volume of data sets which need to summarize certain rules over medical record. This encompasses a novel approach to find the common factors which lead to high risks of diabetes and co morbid conditions associated with diabetes. This performs both association rule mining and association rule summarization techniques with improved classification algorithms.

KEYWORDS: Diabetes, Association, Co-Morbid, Prediction, Split and Merge

INTRODUCTION

Diabetes is a group of metabolic diseases caused by hyperglycemia this is because of defects in insulin secretion, insulin action and both. Next stage chronic hyperglycemia of diabetes is associated with long term damage, dysfunction, and failure of different organs of body, especially the eyes, kidneys, nerves, heart, and blood vessels. This deficiency leads to destruction of the b-cells of the pancreas with consequent insulin deficiency to abnormalities that result in resistance to insulin action and reaction process. The basis of the abnormalities found in carbohydrate, fat, and protein metabolism in diabetes is deficient action of insulin on target tissues. Deficiency of insulin results from inadequate insulin secretion otherwise diminished tissue responses to insulin at the complex pathways of hormonal reaction in the body. Improper insulin secretion and defects in insulin action that is frequently coexist in the same patient and it is often unclear which abnormality is the primary cause of the hyperglycemia. The symptoms of marked hyperglycemia include which includes polyuria, polydipsia.

Classification of Diabetes Mellitus and Other Categories of Glucose Regulation:

Assigning a type of diabetes to an individual often depends on the circumstances present at the time of diagnosis, and many diabetic individuals do not easily fit into a single class. For example, a person with gestational diabetes mellitus may continue to be hyperglycemic after delivery and may be determined to have, in fact, type 2 diabetes. Alternatively, a person who acquires diabetes because of large doses of exogenous steroids...
may become norm glycemic once the glucocorticoids are discontinued, but then may develop diabetes many
years later after recurrent episodes of pancreatitis.

Another example would be a person treated with thyroids who develops diabetes years later. Because
thiazides in themselves seldom cause severe hyperglycemia, such individuals probably have type 2 diabetes that
is exacerbated by the drug. Thus, for the clinician and patient, it is less important to label the particular type of
diabetes than it is to understand the pathogenesis of the hyperglycemia and to treat it effectively.

Abnormality Diagnosis:
In data mining, abnormality detection is the search for data items in a dataset which do not conform to an
expected pattern. The patterns thus detected are called anomalies and often translate to critical and actionable
information in several application domains. Anomalies are also referred to as outliers.

A good definition of an abnormality is as follows “an abnormality is an observation that deviates so much
from other observations as to arouse suspicions that it was caused by a different mechanism”. Distance-based
measures have been used in algorithms to delineate outliers or abnormal records from normal records

Abnormality detection is the process of identifying abnormal pattern from set of objects. Abnormality
detection aims to identify a small group of instances which deviate remarkably from the existing data. A well-
known definition of “abnormality” is given “an observation which deviates so much from other observations as
to arouse suspicions that it was generated by a different mechanism,” which gives the general idea of an
abnormality and motivates many abnormality detection methods.

- Abnormality detection refers to the problem of finding patterns in data that do not conform to expected
  normal behaviour.

- These anomalous patterns are often referred to as outliers, anomalies, discordant observations,
  exceptions, faults, defects, aberrations, noise, errors, damage, surprise, novelty, peculiarities or contaminants in
different application domains.

![Fig. 2: abnormality detection](image)

II. Literature Survey:
A. Image acquisition technique:
With the use of ECG data, abnormal cardiac conditions can be diagnosed. Even though the diagnosis is
simple, but diagnose reports was wrong. Some authors [3] in the literature have proposed image acquisition
techniques to get data from ECG image. Mean while this uses histogram validation for improving the quality of
the image, this has been used to diagnose report automatically. The system [4] renovated with two different
models. The first model is an ontological model and subsequent software system to promote open exchange and
the second one is presentation of ECG data along with the ontology as an aid for the automatic diagnosis of
common cardiac abnormalities. The common method has been used to exchange the visualized result of medical
data.

B. Heart rhythm abnormality finding:
Hidden Markov models (HMM) have been used to inspect ECG waveforms to find abnormal heart rhythm
[5]. The HMM approach combines structural and statistical knowledge of the ECG signal in a single parametric
model. The Model parameters are estimated from training data using an iterative, maximum-likelihood re-
estimation algorithm. However, the mechanisms of detecting cardiac abnormality from ECG waveforms by
applying MLr-E, is successful and the system doesn’t handle P-wave data’s. Therefore, to identify all the
cardiac abnormalities, the presented system in [6] requires hundreds of complex algorithms to be integrated
under one computationally system. Maintaining and updating such a system for every new abnormality is
intrinsically complex. This introduces a problem of finding a simple and fast solution toward heart disease
recognition from compressed ECG that raises alert to the cardiac specialist as soon as a cardiac disease is recognized. In the previous work [7], only performs cardiac abnormality detection with essentially two clusters which are known as normal cluster and abnormal cluster.

C. ECG feature extraction:
Feature Extraction from ECG data [8] plays a major role in detecting most of the cardiac diseases and its risk. The structural information of the ECG signal holds useful information about the nature of cardiac diseases and its abnormalities. The corresponding details are difficult to analyze visually by the humans, thus computer assisted analysis and classification of cardiac diseases can help physicians to monitor cardiac health easily and accurately. Thus, computer-aided automatic diagnosis and classification of cardiac abnormalities is very helpful in health care, which is more helpful in emergency conditions.

III. Proposed System:

The first phase of our proposed work involves the feature extraction process from ECG dataset. The system proposes a new feature extraction method with demographical feature finding. For effective feature extraction the system performs ARHMM model. The following figure 3.0 represents the sample ECG report, which has been taken for experiment. In the ECG report, there are demographical features specified with wave details, using the sample report, the feature has been extracted. The figure 3.0 shows the sample ECG report format, which includes age, gender etc., and the sample ECG digital report has temporal data as x-axis and amplitude as y-axis along with the start, end and peak values.

Fig. 3.0: ECG sample report for feature extraction

The above ECG data has been converted into XML documents, where those XML documents will take into the preprocessing stage. This job completed with the use of the Matlab tool. The initial ECG digital data sample has displayed below (see fig 4.0).
The XML Schema consists of structures for the components of the waveform drawn from the x and y coordinates of the graphical waveform which showed in fig 5.0. The x-axis represents the time interval in seconds while the y-axis represents amplitude in millivolts.

![XML Schema](image)

**Fig. 5.0**: ECG digital report converted to XML

### A. Ahmm:

AutoRegressive Hidden Markov Model (ARHMM) is a combination of autoregressive time series and Hidden Markov Chains (HMC). The Interpretation is generated by a few autoregressive time series while the switches between each autoregressive time series are controlled by a HMC. A time series may sometimes consist of observations generated by different mechanisms at different times. When this happens, the time series observations would act like switching back and forth between couple of distinct states. When changing into a different state, time series may have a significant change in their means or in their frequencies or breadths of their fluctuations. The Autoregressive Hidden Markov model (ARHMM) is often being used to deal with this kind of time series. As indicated by the name, an ARHMM is the combination of an autoregressive time series model and a hidden Markov model. The autoregressive structure admits the existence of dependency amongst time series observations while the hidden Markov chain could capture the probability characteristics of the transitions amongst the underlying states. Actually, ARHMM is also referred as *time series with change in regime (or states)* by the econometricians.

To be more specific, let us see an example of ARHMM. As usual, \( Y = \{Y_1, Y_2... Y_T\} \) denote the observation sequence. Each \( Y_t \) is a observation vector with k component \( Y_t = \{y_1, y_2... y_k\} \).

\( X = \{X_1, X_2...X_T\} \) is a hidden state sequence with N possible states. X is assumed to be a Markov chain with transition matrix \( A = [aij] \) and initial distribution vector \( \pi = \pi_i \).

But it should be mentioned that the ARHMM with demographical for distinct state \( X_t \) could also be developed with more complexity. In such cases, the error term \('t' \) will usually be replaced by \('X_t' \) which depended on the value of current state \( X_t \). E-M algorithm or segmental K-mean algorithms could only lead to a local maximum of the HMM likelihood function. For ARHMM, this is also true.

To get the parameter estimates with a global maximum likelihood, a grid search approach might be used. In grid search approach, the parameter space is seen as a grid with many small cells and all the vertices are used as the initial values of the parameters. Because the parameter space is so big in the case of ARHMM, the grid search method requires considerable computational power which is intractable for practical purposes.

Another notable feature of ARHMM estimation is the high autoregressive coefficients. This is exactly the reason why ARHMM are superior to conventional HMM in this application. Conventional HMM assumes there are independency relations between the observations. But this is rarely the case for time series observations. As in this system, ECG data are collected on a day-by-day bases and apparently the independency assumption is
inappropriate. Comparatively, the autoregressive structure contributes the superiority of ARHMM in a way it prevents the frequent fluctuations of state path. Conventional HMM are very sensitive to the numerical swings of the current ECG and hence mistakes several fluctuations of ECG as the switches of states. While for the same data, ARHMM state path are more stable and close to reality. Using the ARHMM, the ECG data can be diagnosed and reported as graphical/text formats on web browser.

IV. Experiments And Results:
A. Dataset:
   For experiment, our system used dataset from physionet domain and the dataset named as “MIT-BIH Arrhythmia Database” which contains 234 records. The followings are the sample digitized data for each option (see fig 6.0).

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c|c}
\text{Record} & \text{N} & \text{V} & \text{F} & \text{O} & \text{N} & \text{E} & \text{P} & \text{P} & \text{O} \\
\hline
\end{array}
\]

Fig. 6.0: Beat types and its symbol representation

From the above fig 6.0, the beat type has been displayed and that will be extracted as an XML file and applied into the diagnosis process. Then the system finds the abnormality based on the diagnosis rules listed in the HL7 medical device communication standard. The results are converted to output in a browser. By this feature the users can able to see output with abnormality detection in a textual and graphical form without the use of any software.

B. Implementation:
   The D-ECG framework has been developed in C#.net with feature extraction and diagnosis, later this has been presented in a web browser using ASP.net framework.

Fig. 7.0: the result from D-ECG framework
The above fig 7.0 shows the output of the proposed framework, where the graphical and textual data are summarized on the web page. From the above figure, the chart is represented for risk assessment based on the ECG result. From the above experiment, the abnormality has been found in term of percentage. The risk has been calculated for every time interval and it also displays the overall finding with demographical features.

**Conclusion:**

The developed D-ECG framework provides a graphical construction for the demonstration and sharing framework of ECG data so that it can be made readily accessible for presentation or screening on a large amount of computing environments. Using the ECG digital data as input the user can diagnosis abnormalities of cardiac conditions. These resources are based on the HL7 standard and AHMM model. The improved ontology structure used for data representation and thus provides an easy presentable text format for the user after successful risk analysis via web browser.

**REFERENCES**